Detection and Segmentation of Underwater Archaeological Sites Surveyed with Stereo-Vision Platforms

Katherine A. Skinner
Robotics Program
University of Michigan
Ann Arbor, Michigan 48109
Email: kskin@umich.edu

Matthew Johnson-Roberson
Department of Naval Architecture and Marine Engineering
University of Michigan
Ann Arbor, Michigan 48109
Email: mattjr@umich.edu

***Abstract*—This paper proposes a method for automating detection and segmentation of archaeological structures in underwater environments. Underwater archaeologists have recently taken advantage of robotic or diver-operated stereo-vision platforms to survey and map submerged archaeological sites. From the acquired stereo images, 3D reconstruction can be performed to produce high-resolution photo-mosaic maps that are metrically accurate and contain information about depth. Archaeologists can then use these maps to manually outline or sketch features of interest, such as building plans of a submerged city. These features often contain large rocks that serve as the foundation to buildings and are arranged in patterns and geometric shapes that are characteristic of human-made structures. Our proposed method first detects these large rocks based on texture and depth information. Next, we exploit the characteristic geometry of human-made structures to identify foundation rocks arranged along lines to form walls. Then we propose to optimize the outlines of these walls by using the gradient of depth to seek the local minimum of the height from the seafloor to identify the ground plane at the base of the rocks. Finally, we output contours as geo-referenced layers for geographic information system (GIS) and architectural planning software. Experiments are based on a 2010 stereo reconstruction survey of Pavlopetri, a submerged city off the coast of Greece. The results provide a proof-of-concept for automating extraction of archaeological structure in underwater environments to produce geo-referenced contours for further analysis by underwater archaeologists.**

I. Introduction

The field of underwater archaeology has seen recent advances in photogrammetric survey techniques that take advantage of robotic or diver operated stereo-vision platforms to map submerged archaeological sites. The advantage of these platforms is that large areas can be mapped more efficiently than with traditional methods that require divers to manually measure submerged structures. The stereo images gathered can be used to perform 3D reconstructions of submerged sites, resulting in high-resolution photo-mosaic maps with recovered depth data. These metrically accurate reconstructions allow for further mapping and measurement of the submerged site without the need for repeated dives or vehicle deployments [1]. Using these reconstructions, archaeologists typically trace or manually sketch city or building plans of the site. An example of such an archaeological plan produced in [1] appears in Fig. 1.

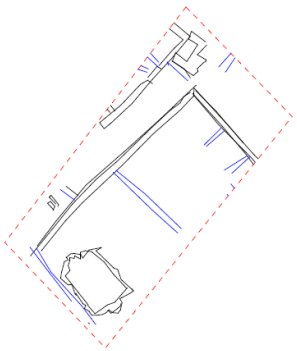


Fig. 1. Hand-drawn building plans produced in [1] during a survey of Pavlopetri, a submerged city off the coast of Greece.

The features of interest to underwater archaeologists often contain large rocks distinct from the surrounding environment that serve as the foundation to buildings and are arranged in geometric shapes that are characteristic of human-made structures. In this paper, we propose a method for exploiting these characteristics to automate detection and segmentation of structural features within these sites, resulting in vectorized contours of submerged archaeological structures. The proposed method is developed and tested with photo-mosaics and depth information obtained through stereo reconstruction of data gathered with a diver rig during a 2010 survey of Pavlopetri, a submerged city off the coast of Greece [2]. The output is a geo-referenced contour of detected archaeological structures that can be layered on top of the original photo-mosaic map for use in geographic information system (GIS) or architectural planning software.

This paper is organized as follows: Section II provides background and prior work on detection and segmentation of features in archaeological sites. Section III develops our proposed technical method for the specific application of underwater archaeological sites. Section IV presents experi-

mental results based on a 2010 survey of the submerged city of Pavlopetri, Greece, and Section V gives a discussion of these results. Finally, Section VI concludes with suggestions for future work.

II. Background

Prior work on automating detection of structures in archaeological sites has mainly been applied to large-area terrestrial sites mapped with remote sensing platforms. Lambers and Zingman used large-area satellite images to detect regions with a high likelihood of containing features of interest to archaeologists based on texture contrast and geometric cues such as straight line walls meeting at corners [3] [4]. D'Orazio et al. [5] developed a Modified Variance Analysis method to detect lines characteristic of archaeological sites from satellite images. Luo et al. [6] extract circular structures from Google Earth images to point out qanat shafts for agriculturalists. These methods take advantage of the geometric properties characteristic of archaeological sites, but they do not incorporate depth information and are not applied to the same spatial scales that we are working with.

More generally, rock detection has been studied for autonomous planetary science [7]. Golumbek [8] exploited shadows produced by rock formations to identify and characterize rocks on Mars. Dunlop et al. [9] and Niekum and Wettergreen [10] both integrated multiple properties such as texture, size, and shape to detect and segment rocks on Mars. However, the terrain of Mars features a smooth sand environment with rocks that cast consistent shadows, whereas underwater environments are often rocky or coarse and have directionally dependent shadowing effects [11].

In underwater environments, relevant work involves classification of sea-bottom terrain [12], allowing for detection of features such as rocks, which are useful for identifying building foundations at archaeological sites. Methods specifically applied to detection of submerged archaeological objects were developed by Moroni et al. [13], based on optical and acoustic data. Douillard et al. performed segmentation of undersea terrain on point clouds obtained through 3D laser line imaging [14] [15]. In contrast, the work presented in this paper focuses on optically derived measures gathered with a stereo-vision platform.

III. Methodology

Our proposed method takes as input a high-resolution photo-mosaic map reconstructed from a stereo-vision survey, as well as depth data recovered from the 3D stereo reconstruction. Figure 2 shows a high-resolution photo-mosaic map of a rectangular room reconstructed from stereo-vision imagery gathered during a survey of Pavlopetri, Greece in 2010. Figure 3 shows a depth relief recovered through the same 3D reconstruction. Our proposed method then has several steps described as follows:

Fig. 2. High-resolution photo-mosaic map of a rectangular room reconstructed from stereo-vision imagery gathered during a survey of Pavlopetri, Greece in 2010.

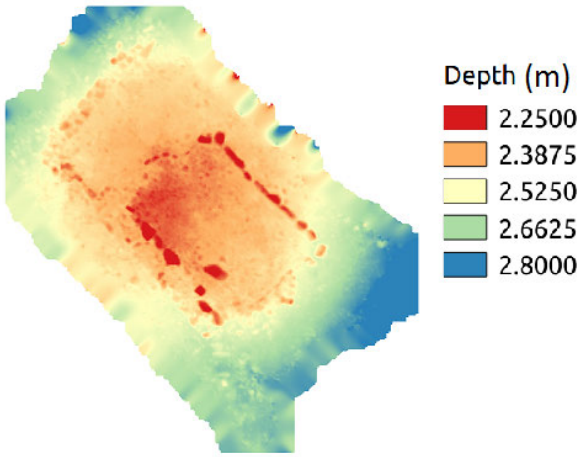


Fig. 3. Depth relief recovered through a 3D reconstruction from stereo images.

A. Initial Segmentation

To initially segment the images, we compute superpixels using the method of Simple Linear Iterative Clustering (SLIC) [16] [17]. Superpixels result in an over-segmentation of the image that provides initial segment contours. An example of a typical superpixel segmentation of the input photo-mosaic is shown in Fig. 4. This is a good initial segmentation, as the contours are closed and the scale of each segment is on the same scale as that of the large foundation rocks. However, the contours follow lines of contrast and do not always produce accurate region boundaries. For example, when there is dark growth covering part of a rock, the superpixel segments the rock into light and dark sections of the same rock. Therefore, optimization of these contours is required to achieve a more accurate segmentation.

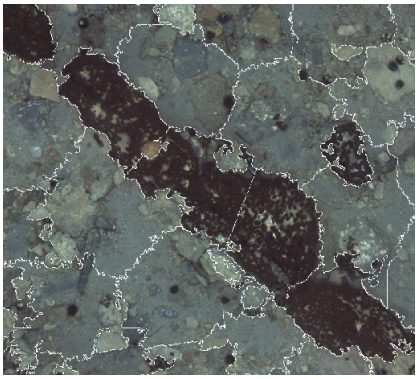


Fig. 4. Zoomed in section showing an example of a typical superpixel segmentation of the input photo-mosaic to show the size and shape of initial contours.

B. Texture Classification

Next we use a supervised texture classification method to classify each superpixel as either a rock or non-rock segment using the texture classification method presented by Varma and Zisserman [18], which is based on the rotationally invariant Maximum Response-8 (MR8) filter bank [19] and textons, which are defined by Leung and Malik [20] as the centers of clustered filter responses. This method has two stages: the learning stage and the classification stage.

As outlined in Fig. 5, the first step of the learning stage is to create a texton dictionary of different textures found among both classes. To create the dictionary, input images are convolved with the MR8 filter bank, which includes six orientations of edge and bar filters at three different scales, with the maximum response selected at each scale, a Gaussian filter, and a Laplacian of Gaussian filter. This produces 8 filter responses per pixel. Using images across both rock and non-rock classes, pixel responses are aggregated and clustered using k-means clustering [21] to obtain a dictionary of 16 8-dimensional textons based on cluster centers.

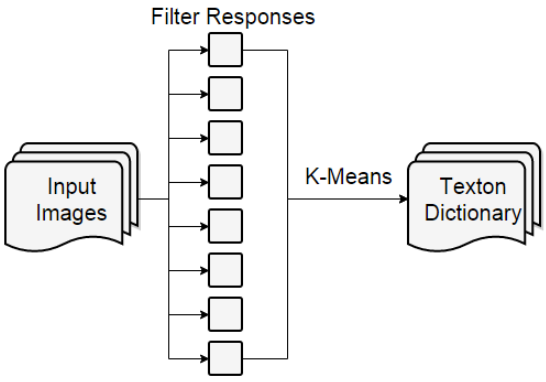


Fig. 5. Outline of the first step in the learning stage for texture classification. Input images are convolved with the MR8 filter bank. Eight-dimensional filter responses for each pixel are aggregated and clustered using k-means. The cluster centers of filter responses make up a texton dictionary of textures from both classes.

The next step is to learn models of rock and non-rock classes based on labeled training images and the generated texton dictionary. As outlined in Fig. 6, for each training image, a histogram of texton frequencies of the training image is computed by comparing each 8-dimensional pixel response to a texton in the dictionary according to Euclidean distance nearest neighbor, and adding a count for each pixel nearest to each texton in the dictionary. Note since training and testing images are different sizes, histograms are normalized to sum to 1. The resulting histogram of each training image provides a model for the labeled image class.

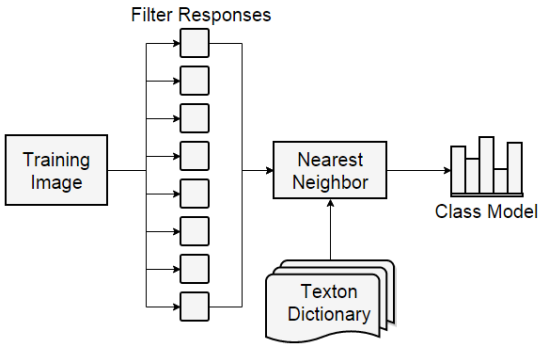


Fig. 6. Outline of the second step in the learning stage for texture classification. The filtered responses of each pixel in a training image are compared to textons in the texton dictionary. A histogram is created from counts of the closest texton to each pixel and the histogram is normalized to sum to 1. The resulting histogram acts as a single model for the labeled class of the training image.

During the classification stage, a histogram is similarly computed for each testing image, and compared to the class models produced through training based on Euclidean distance nearest neighbor in order to determine which model the test image is closest to and, subsequently, which class the test image belongs to.

We trained our classifier on 80 rock and 80 non-rock images, and achieved 90% accuracy for classifying rocks and 88% accuracy for classifying non-rocks on a test set of 60 labeled images per class.

C. Integrating Depth Information

As another metric of rock classification, we note that the depth information provides an important additional cue of structure along the seafloor. Figure 7 shows the depth profile of a wall segment, where it is clear that the larger rocks forming the wall stand out against the ground plane. However, absolute depth cannot be used directly since the elevation of the ground plane changes over large areas; instead we note that the gradient of the depth, shown in Fig. 8 also distinguishes the wall segment from the ground plane, so we set a threshold on the average depth gradient across a segment required to classify the segment as a structurally significant rock. Figure 9 shows the segmentation of significant rocks according to both texture and depth gradient for a submerged rectangular room surveyed at Pavlopetri. The combination of visual and depth

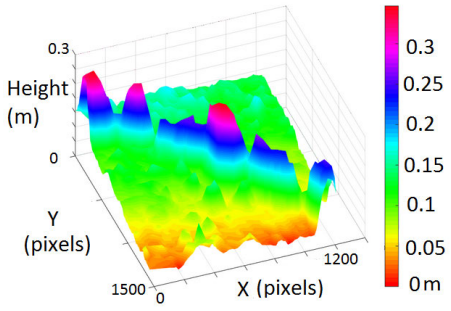


Fig. 7. The depth contour here highlights large rocks that form a segment of a wall from the rectangular room shown in Fig. 2.

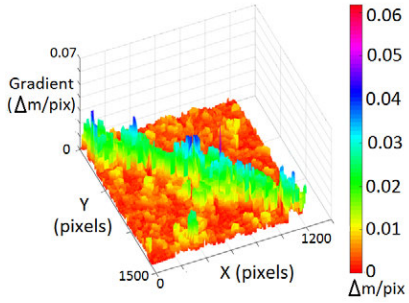


Fig. 8. Taking the gradient of the depth over the same segment shown in Fig. 7 can also provide an important cue for identifying foundation rocks.

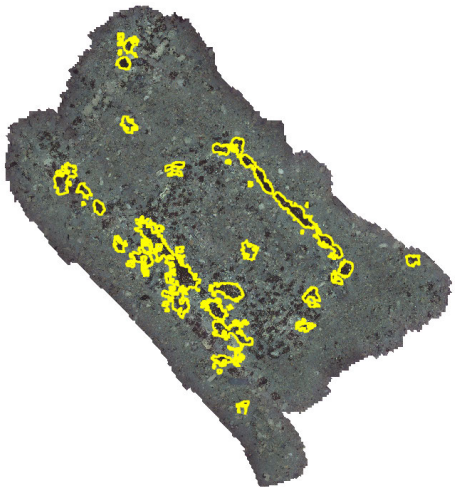


Fig. 9. Contours of all superpixels labeled as rocks using texture classification integrated with depth information.

modality provides a distinction between large rocks that have potential to serve as building foundation, and rocky or coarse regions of the seafloor.

D. Integrating Geometry

To identify potential archaeological features among detected foundation rocks, we note that man-made structures often have distinct geometric properties. For example, walls fall in straight lines and meet at corners. To exploit this straight-line geometry, we employ Random Sample Consensus (RANSAC) [22] to identify lines of best-fit among segments classified as rocks. Figure 10 shows three best-fit lines detected in the rectangular room. We traverse these lines to identify full structures, or clusters of rocks, by including all identified rocks along the line. From each rock on the line, we search connected superpixels to attach to the cluster, until we reach a superpixel not identified as a rock. We eliminate outliers by setting a distance threshold between consecutive rocks in the cluster.

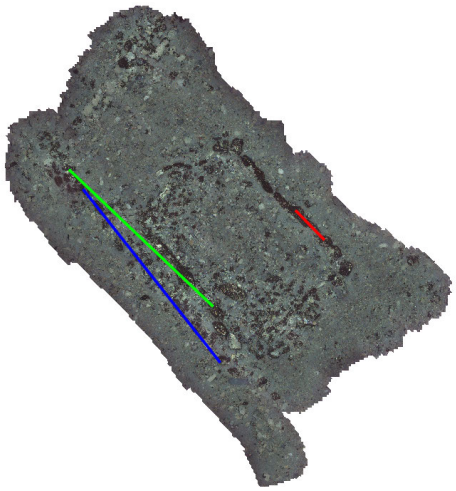


Fig. 10. Best-fit lines computed by RANSAC given detected rocks.

E. Refining Contours

Given a single structure, or cluster of rocks, we seek to improve the contours around rocks from the initial superpixel segmentation in order to encompass the entire rock. Figure 11 shows a sample depth profile of a single rock, from which we can observe that the base of the rock tends to lie at a local minimum height from the seafloor. To identify the base, we search neighboring pixels to each pixel in the initial contour and perform gradient descent along neighboring pixels according to the height from the seafloor. We terminate when gradient descent reaches a local minimum or if the depth gradient decreases below a threshold indicating that the ground plane has been reached. This enables optimization of the initial rock contours to more accurately segment foundation rocks from the surrounding seafloor.

The output contour is exported to a GIS program or architectural planning software and can be displayed on top of the original photo-mosaic or on its own as a geo-referenced contour layer.

IV. Experimental Results

We tested the above proposed method with real data from a 2010 field survey of Pavlopetri, Greece. For this dataset,

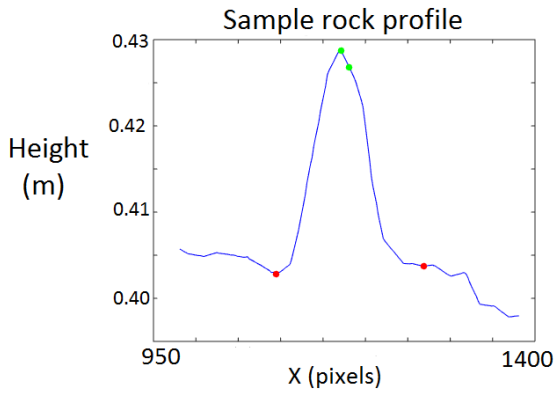


Fig. 11. Sample rock profile with start (green) and end (red) points of gradient descent.

stereo images were gathered using a stereo-vision diver rig platform. Stereo reconstruction was performed to obtain a high-resolution photo-mosaic image of the survey area with recovered depth data. Here we show results for a survey of a rectangular room at the site.

Figure 12 shows the final clustered segmentation of independent structures, or walls, found along the best-fit lines shown in Fig. 10, prior to gradient descent, where each colored cluster is identified as a separate wall. Figure 13 shows the final wall contours after gradient descent is performed, showing that bases of the rocks can be captured by seeking the minimum height from the bottom near each rock region. Figures 14 and 15 show these results exported as vectorized contours for visualization as geo-referenced layers in GIS software.

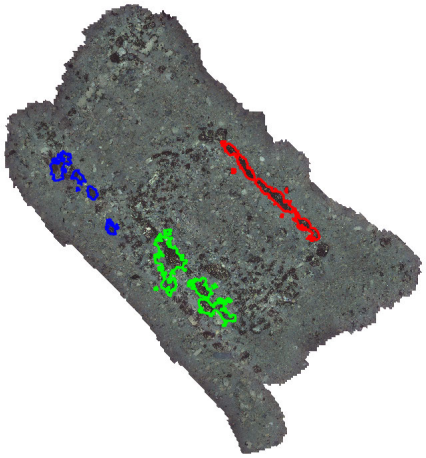


Fig. 12. Final clustered segmentation of independent structures, or walls, prior to gradient descent, where each colored cluster is identified as a separate wall.

V. Discussion

These results validate our methods as a proof-of-concept for detecting archaeological structure in underwater environments.

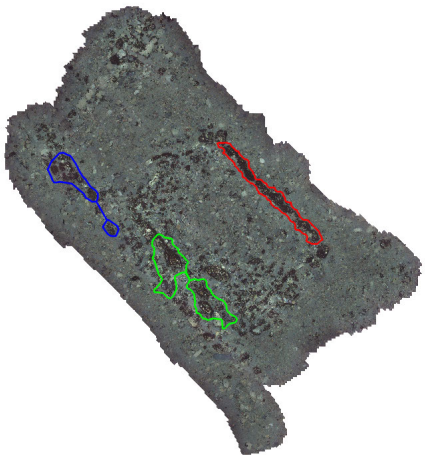


Fig. 13. Final wall contours after gradient descent is performed, where each colored cluster is identified as a separate wall.

However, there are some failure modes that could be improved for better results. One failure mode of the initial rock detection occurs in the case of lighter rocks, which are often mislabeled as non-rock regions. The texture training method we use is dependent on the training set and the diversity of the texture such that the lighter rocks may resemble smooth or sandy non-rock regions. To improve this, we propose adding additional measures such as rock size or shape to identify potential rock segments. Other methods of unsupervised learning may also be explored to better classify rock regions.

Our methods do show some robustness to poorly fit lines found along rock segments. For instance, the green line in Fig. 10 extends past the lower wall. As we assume that rocks belonging to the same wall are typically found close together, we require that rocks along the wall cannot be a certain distance apart from each other. Therefore, we can successfully eliminate rocks at the top segment of the line from being included in the wall. Future work will focus on increasing this robustness for other cases such as neighboring walls. As also seen along the green wall, a rock is included in the detected wall of the interior room that most likely belongs to the external wall, as the rocks of each wall are placed close together.

To examine results of gradient descent, Fig. 16a provides the initial contours of an identified rock region prior to gradient descent. This shows a failure mode of the initial contour scheme such that the superpixels are split across the same rock, dividing segments between the lighter and darker portion of the rock. Figure 16b shows a sample result of the same region after gradient descent is used to expand contours from the initial superpixels. This contour more completely captures the base of the rock as desired. Note that superpixels identified as rocks that contain below a set number of pixels are eliminated. Additionally, the gradient descent does not provide a smooth contour according to the depth, so the resulting contours are

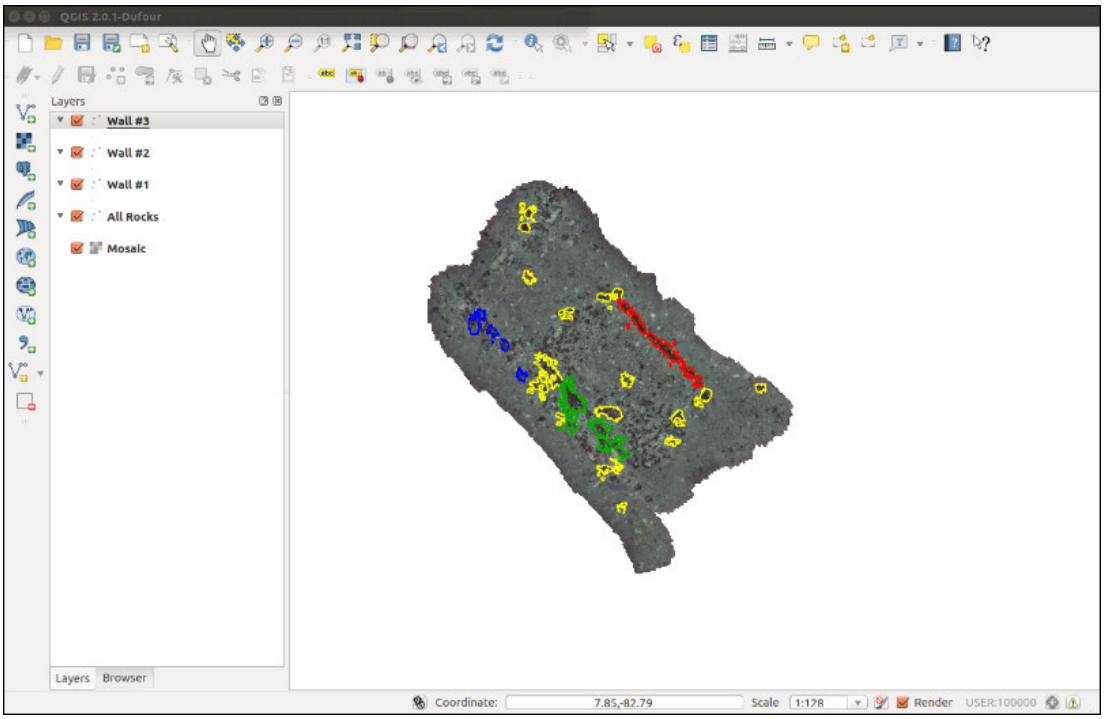


Fig. 14. Contours exported to GIS software and displayed as individual geo-referenced layers on top of the input photo-mosaic. The yellow layer represents all detected rocks with potential to be foundation rocks. The red, green, and blue layers are individual layers representing detected structures, or walls.

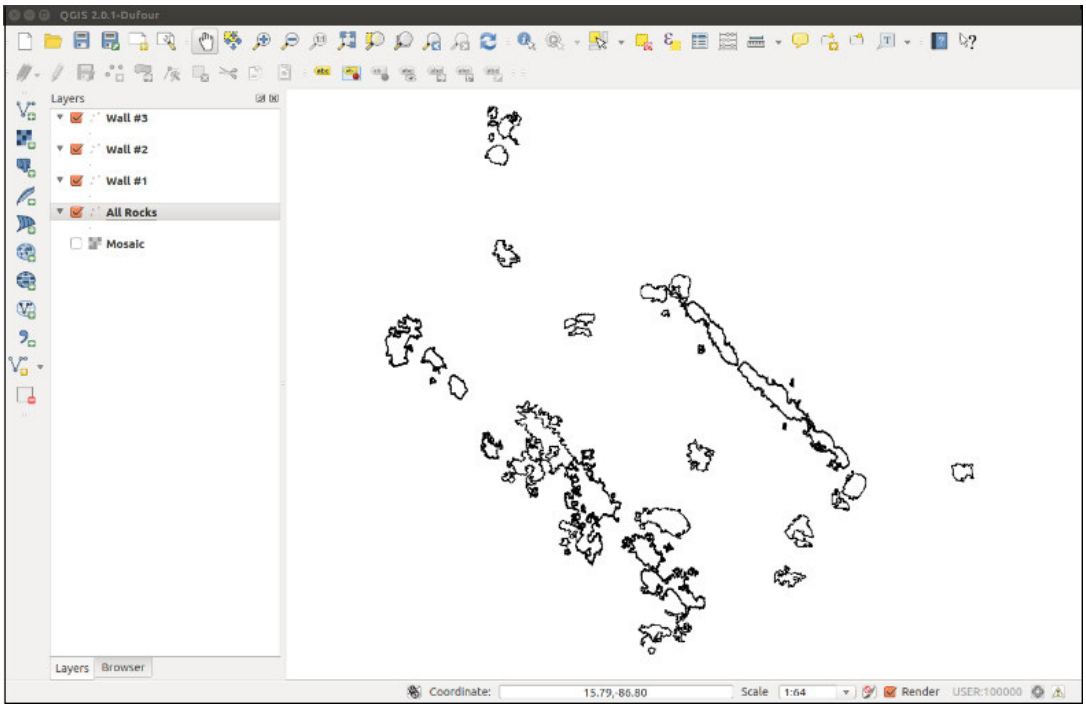
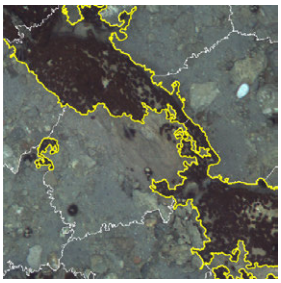
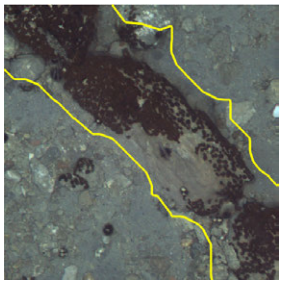


Fig. 15. Contours of all detected rocks with potential to be foundation rocks are exported to GIS software and displayed as an individual geo-referenced layer.



(a) Initial contours prior to gradient descent.



(b) Expanded contours based on gradient descent.

Fig. 16. Sample result of gradient descent showing expanded contours that encompass the base of the rock.

based on a non-convex boundary around the points reached by gradient descent, which provides a smoothed boundary.

VI. Conclusions & Future Work

While the expert knowledge of underwater archaeologists in interpreting archaeological structure is essential, tasks such as manually tracing this structure, or searching large areas to detect features of interest is often tedious and time consuming. Therefore, it is desirable to automate these tasks. These results validate the proposed method as a proof-of-concept for automating extraction of structure in underwater archaeological sites from input photo-mosaics and depth information to produce geo-referenced contours to be used as layers in GIS and architectural planning software. Future work will focus on optimizing rock contours, increasing robustness, and analyzing computational complexity to increase efficiency of the system. Ultimately, this could lead to the development of a complete tool to assist underwater archaeologists in their work.

Acknowledgment

The authors would like to thank the Australian Centre for Field Robotics for gathering the data, as well as our colleague Jon Henderson for archaeological input. The work presented has been partially funded by the National Oceanic and Atmospheric Administration under grant number NA14OAR40110265.

References

[1] I. Mahon, O. Pizarro, M. Johnson-Roberson, A. Friedman, S. Williams, and J. Henderson, "Reconstructing pavlopetri: Mapping the world's oldest submerged town using stereo-vision," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 2315–2321.

[2] J. Henderson, O. Pizarro, M. Johnson-Roberson, and I. Mahon, "Mapping submerged archaeological sites using stereo-vision photogrammetry," *International Journal of Nautical Archaeology*, vol. 42, no. 2, pp. 243–256, 2013.

[3] K. Lambers and I. Zingman, "Towards detection of archaeological objects in high-resolution remotely sensed images : the silvretta case study," in *Archaeology in the Digital Era : Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012*, P. Verhagen, Ed., 2013, pp. 781–791.

[4] ——, "Texture segmentation as a first step towards archaeological object detection in high-resolution satellite images of the silvretta alps," *Archaeological Prosepection : Proceedings of the 10th International Conference - Vienna May 29th - June 2nd 2013*, pp. 327–329, 2013.

[5] T. D'Orazio, P. Da Pelo, R. Marani, and C. Guaragnella, "Automated extraction of archaeological traces by a modified variance analysis," *Remote Sensing*, vol. 7, no. 4, pp. 3565–3587, 2015.

[6] L. Luo, X. Wang, H. Guo, C. Liu, J. Liu, L. Li, X. Du, and G. Qian, "Automated extraction of the archaeological tops of qanat shafts from vhr imagery in google earth," *Remote Sensing*, vol. 6, no. 12, pp. 11 956–11 976, 2014.

[7] D. R. Thompson and R. Castano, "Performance comparison of rock detection algorithms for autonomous planetary geology," in *Aerospace Conference, 2007 IEEE.* IEEE, 2007, pp. 1–9.

[8] M. Golombek, A. Huertas, D. Kipp, and F. Calef, "Detection and characterization of rocks and rock size-frequency distributions at the final four mars science laboratory landing sites," *Mars*, vol. 7, pp. 1–22, 2012.

[9] H. Dunlop, D. R. Thompson, and D. Wettergreen, "Multi-scale features for detection and segmentation of rocks in mars images," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–7.

[10] S. Niekum and D. Wettergreen, "Reliable rock detection and classification for autonomous science," *Carnegie Mellon University Thesis*, 2005.

[11] S. Q. DUNTLEY, "Light in the sea," *J. Opt. Soc. Am.*, vol. 53, no. 2, pp. 214–233, Feb 1963.

[12] A. Martin, G. Sévellec, and I. Leblond, "Characteristics versus decisions fusion for sea-bottom characterization," in *Proceedings of the Colloque Caractérisation in situ Des Fonds Marins, Oct*, 2004, pp. 21–22.

[13] D. Moroni, M. A. Pascali, M. Reggiannini, and O. Salvetti, "Underwater scene understanding by optical and acoustic data integration," in *Proceedings of Meetings on Acoustics*, vol. 17, no. 1. Acoustical Society of America, 2014, p. 070085.

[14] B. Douillard, N. Nourani-Vatani, M. Johnson-Roberson, S. Williams, C. Roman, O. Pizarro, I. Vaughn, and G. Inglis, "Fft-based terrain segmentation for underwater mapping," in *Robotics: Science and Systems (RSS)*, 2012.

[15] B. Douillard, N. Nourani-Vatani, M. Johnson-Roberson, O. Pizarro, S. Williams, C. Roman, and I. Vaughn, "Frequency-based underwater terrain segmentation," *Autonomous Robots*, vol. 35, no. 4, pp. 255–269, 2013.

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," Tech. Rep., 2010.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.

[18] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.

[19] ——, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, vol. 3. Springer-Verlag, May 2002, pp. 255–271.

[20] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision*, vol. 43, no. 1, pp. 29–44, Jun. 2001.

[21] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[22] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.